

DOCUMENT RESUME

ED 198 166

TM 810 133

AUTHOR Reid, Neil A.
 TITLE Research on Educational Measurement and Testing in New Zealand.
 INSTITUTION New Zealand Council for Educational Research, Wellington.
 PUB DATE 79
 NOTE 28p.; Paper presented at the National Conference of the New Zealand Association for Research in Education (1st, Wellington, New Zealand, December 7-10, 1979).
 EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS *Educational Research; *Educational Testing; *Foreign Countries
 IDENTIFIERS *New Zealand; New Zealand Council for Educational Research

ABSTRACT

This review of the research on educational measurement and testing in New Zealand uses the institutional setting of the research as its framework. Following an historical overview dating back to the work of Norman McKenzie in 1927, sections of the review deal with research conducted by: (1) university staff; (2) university students; (3) Department of Education; (4) teachers in the schools; (5) teachers' colleges; (5) armed services; (6) other agencies; and (7) New Zealand Council for Educational Research. Despite New Zealand's remoteness and isolation from world centers of test development and research, the country has managed to remain well informed of overseas developments largely through articles in scholarly journals, reports disseminated by large research organizations, and through active participation by New Zealand researchers on study or training leave with overseas institutions. With the growth of a greater research capability within the Department of Education and the establishment of the Test Development Division by the New Zealand Council for Educational Research, tests or assessment devices designed to fulfill particular functions have been developed specifically for New Zealand conditions.
 (Author/RL)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

NZARE

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE-
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Research on Educational Measurement and Testing in New Zealand

Neil A. Reid
Test Development Division
New Zealand Council for Educational Research

To provide a coherent, comprehensive and approximately truthful account of research activity on educational measurement and testing in this country over the past 30 years is by no means easy, especially as so little appears to have been recorded for posterity. Tracking down the elusive facts I have found to be a major obstacle. I do not have all of them by any means so that the picture of research I project may have blank spots; it may be distorted because of my particular viewpoint, and possibly out of focus too, depending on your own interpretation of what should be highlighted.

I have begun by limiting the topic to educational measurement and testing. I want to further define the field by making a distinction between research which simply uses tests and other forms of measurement (which would embrace much educational research of an empirical nature) and research which addresses problems of educational measurement, by developing, modifying or adapting tests and measures, evaluating appropriate uses of measurement techniques (both old and new) or illuminating aspects of the administration and interpretation of tests. Hence, to take two recent pieces of research, I would not consider the survey, *Educational Standards in State Schools* (Department of Education, 1978) within the ambit of my definition whereas the development of more appropriate measures of mathematical competence in the Canterbury-Westland scheme (Kay, 1978) is included. Administering a questionnaire,

A paper presented at *Research in Education in New Zealand : A Balance Sheet*, First National Conference of the New Zealand Association for Research in Education (NZARE), Victoria University of Wellington, 7-10 December 1979.

rating scales and achievement tests devised by somebody else and using a sampling design, administration procedures, and data analyses provided by some other institution as in the IEA studies (the current Mathematics study excepted) is not, in my book, New Zealand research on educational measurement and testing (Phillipps, 1976). But taking the *Otis Tests of Mental Ability* and checking on the gradient of item difficulties for a sample of intermediate school pupils quite definitely is (Arvidson, 1974). I trust my distinction is clear because I have applied it rigorously in analysing the few reports and lists of research studies available in this country.

In preparing this review I have used the institutional setting of the research in an endeavour to provide a suitable framework. While this approach works tolerably well, there are limitations and inevitable overlaps with collaborative and interdisciplinary research. No attempt is made to provide detail of studies; the prime objective has been to present a broad perspective.

Historical Overview*

Despite their isolation and remoteness from the centres of test development and research, educators in New Zealand were experimenting with various kinds of tests and measures, notably those of Binet, Ballard and Terman, during the early part of this century. The name of Norman McKenzie is associated with this pioneer work. In 1921 McKenzie went to Canada for training in research and test development. Later, as senior inspector for Taranaki Province, he conducted surveys in 1927 and again in 1929 using such measures as the *Haggerty Reading Examination* and *Test of Educable Capacity*, *Ayre's Scales* and the *Courtis General Test*; tests which he revised for use in New Zealand. The published findings of these Taranaki school surveys, released in 1930, may be regarded as this country's first research monograph. Concurrently, D.G. Ball, Inspector of Native Schools, employed McKenzie's 1929 test battery with 1000 Maori pupils in 92 native schools and reported his findings in the *New Zealand Gazette*, December, 1930.

In this same decade the universities took the initiative, introducing the country to new ways of assessing or diagnosing aptitude and achievement. Although it is difficult to tease out the educational from the more clinically

* This section draws heavily on *Fifty Years of Test Supply - The New Zealand Experience*, (Watson, 1979), a paper presented at the 50th Jubilee Conference, New Zealand Psychological Society, Massey University.

oriented work, men such as Hunter, Gould and Shelley (N.Z.'s first professor of education) were experimenting with tests like those from the *NIIP* and the *Army Beta* in a variety of settings: schools, prisons and industry. Shelley, at Canterbury, established a clinic in 1923, with one of its functions being to undertake educational and vocational testing of children and he appointed a young man fresh from teachers' college, C.E. Beeby, as his assistant. Beeby later became the University's first director of educational and psychological laboratories.

Another name from that era is that of Winifred Valentine, who, after two years abroad studying the mentally retarded, took up an appointment in 1925 to assess children for special classes throughout New Zealand. She worked mainly with the Binet but also with other achievement and performance measures.

Contemporaneously, a number of New Zealand schools, such as Kowhai Junior High School in Auckland, had developed relatively elaborate testing programmes using measures like the *Akarana Test*, the *Pomare Group Test of Intelligence*, both home-grown tests, and other aptitude and achievement tests for guidance and classification purposes.

Extensive studies were also being undertaken in the mid-20s by the Department of Education in preparing the *Terman Intelligence Tests* for use in New Zealand.

The first test to be standardized for use in New Zealand appears to have been Ria MacAlister's *One Minute Reading Test* for children aged 6-11; the work being undertaken in 1928 or 1929.

Into the 1930s and the New Zealand Council for Educational Research became involved in research on educational measurement and testing with the standardization of two measures of intelligence: the well-known *Otis Self-Administering Tests of Mental Ability* and an *ACER Non-verbal Test*. It is interesting to note that the New Zealand standardizations were planned to be conducted simultaneously with comparable standardizations across the Tasman (McIntyre, 1938; Redmond and Davies, 1940). Regrettably, NZCER's plans for comprehensive test development and validation programmes were disrupted by war and the energies of its small staff were directed elsewhere.

Six months before the war ended, NZCER sponsored a national conference devoted exclusively to questions of testing; the first and only conference of its kind to date. The report of the conference and the developments which followed confirm that this meeting was a key event providing guidelines for New Zealand educators and researchers for the following 20 years.

It is interesting to note also, the calibre of the university personnel teaching at this time in the field of educational psychology, tests and measurement - Winterbourn, Parkyn, Tizard, Betty Bernadelli, Beaglehole and others who went on to distinguished academic careers.

During the 1940s a group of graduate students at the University of Auckland, under the leadership of Professor A.B. Fitt, assembled data in a series of theses to modify the *Stanford-Binet Scale* (Form L) for New Zealand (Fitt, 1952). Their work was supplemented by related research projects at Canterbury under Ralph Winterbourn and at Victoria under Colin Bailey and Arthur Fieldhouse. This exercise represents one of the few examples of collaboration among universities in reviewing the suitability of a widely-used test (NZCER, 1948).

Fieldhouse, at Victoria University, also began work in 1946 to adapt a number of achievement measures for use in New Zealand schools. After appraising the reading tests of Burt, Ballard and Schonell he opted instead to prepare the *Oral Word Reading Test* (Fieldhouse, 1952). Under NZCER's sponsorship, Fieldhouse continued his work into the fifties and standardized the ACER reading battery : *Work Knowledge, Speed of Reading and Reading for Meaning* (NZCER, 1956) and the *ACER : Arithmetic Test - Form C* (NZCER, 1956). These tests, incidentally, were not adapted in any way but were straightforward standardization exercises.

Apart from E.W. Seville's development of a series of diagnostic arithmetic tests, which proved very popular with teachers, the work in the area of educational measurement and testing continued to be almost exclusively concerned with the adaptation and/or norming of overseas tests for New Zealand conditions. In the early 1950s, for example, Cyril Rogers, with assistance from students engaged on theses for the University of Auckland, set out to standardize Thurstone's *Primary Mental Abilities Test* on a sample of 11 to 17-year-olds in the province of Auckland (Rogers, 1956); Dick Seddon at Otago University prepared Otago norms for the *Schonell Diagnostic Arithmetic Tests* in 1952; and Peter Freyberg assembled norms for a Manawatu sample for Schonell's *Silent Reading Test A* in 1957.

In the early sixties the Commission on Education (Currie Commission) made a number of recommendations which had a profound impact on work in the testing field. Most notably the Commission recommended the development of standardized tests of attainment in the basic subjects to be administered at regular intervals during a child's school career. (These 'checkpoint' proposals were subsequently modified; see *Education*, October, 1963). As events transpired, the Government formally requested in 1965 that NZCER set up a separate division to undertake the task. The work of this division is discussed below.

University Research - Staff

One could reasonably expect that investigations in the area of educational measurement and testing might be conducted by university staff involved with the teaching of educational measurement courses or the training of students in the administration and interpretation of tests. A search of the university calendars and departmental annual reports from 1970 on, reveals that this is indeed the case, but that the output is somewhat meagre. Staff of our various universities, so far as one can tell from the titles of articles and the type of journal which publishes this kind of research, together averaged fewer than two articles per year and, as far as can be ascertained, in some years nothing at all in the area of tests and measurement was published by a New Zealand university staff member. Only ten or so people have published research during the 9-year period surveyed and three lecturers from three different universities have been responsible for the majority of the publications. The diversity of research completed by university staff is illustrated by the following examples of published findings: *An Objective Method of Defining the Region of the Pass/Fail Cutting Point in University Examinations* (Marsh, 1969), *Some Effects of Testing Procedure on Divergent Thinking* (Nicholls, 1971), *The WISC-R : A New Zealand Study of Norms and Validity* (Tuck, Hanson and Zimmerman, 1975), *The Stability of Scores and Difference Scores on the Progressive Achievement Tests of Listening and Reading Comprehension* (Hughes and Tuck, 1978).

With so few published studies over a wide range of topics within the broad field, it is obviously impossible to detect any trends;

aspects of the field of particular interest at a given point in time have been researched by an individual and in some cases the researcher has moved from one topic to another over the years; in others several studies explore one aspect, or it may be that one study spawns several articles.

But again, this reporting of available 'facts', largely publications, does not disclose the full extent of experimentation and research in this area. Undoubtedly, university staff have conducted small-scale studies on a wide variety of problems. These have never been published; very often they have not even been written up in report form, so that unless one happens to learn by way of personal communication of some kind in the course of one's own investigations, the scope and frequency of this kind of research is not disclosed.

Major test construction exercises have been conducted in recent years in at least two universities, Auckland and Canterbury. At Auckland, Marie Clay (Clay, 1972; Clay *et al.*, 1976) has produced two measures in the language area and Tony McNaughton (McNaughton, 1974) has been responsible for the development of a test of social studies.

Some years ago less successful efforts were made at Canterbury. The ill-fated *Canterbury Reasoning Test* and the *Tasman Tests* were developed and published (Shouksmith, 1960; 1964) but failed, for a variety of reasons, to capture a large test-user audience; Keeling and Seddon worked for some time on a test battery for Form III entrants (Keeling and Seddon, 1969) but the projected full battery has never been completed. Perhaps also worthy of mention is Jeanne Perry's abortive attempt to construct a performance test for Cook Islands children (NZCER, 1972).

Over 20 years ago, Gordon Arvidson (Arvidson, 1959) developed an entrance test known as the *General Ability Test* while on the staff of Victoria University, a measure which is still used with satisfaction by a number of secondary schools in the southern part of the North Island and elsewhere.

As an example of interdisciplinary research involving a sizeable educational measurement and testing component, Otago University's Dunedin Multidisciplinary Child Development Study is the obvious project upon which to focus. Phil Silva, Jennifer Bradshaw and others have developed several measures, most notably *The Dunedin Reading Test*, and investigated other tests such as the *Reynell Developmental Language Scale*, in the course of their longitudinal research of children 'at risk' (Silva, 1978; Bradshaw, 1978).

Presently, several New Zealand universities have small specialist units, and while each has a slightly different orientation, all have as one of their stated functions the improvement of teaching and assessment techniques of university staff. John Jones, Higher Education Research, Auckland University, John Clift, University Teaching and Research Centre at Victoria, Tuan Emery at the Educational Research and Advisory Unit, Canterbury University, and Terry Crooks, Higher Education Research Advisory Centre, Otago University, have all been involved in evaluation exercises which could legitimately be considered a form of research in the tests and measurement field. Publications for general consumption are not numerous, as yet, but it should be remembered that staff of these units have a rather special relationship with their clients which must be considered.

University Research - Students

Unpublished degree and diploma studies categorized under the heading of Tests and Testing in listings which have appeared irregularly in the *New Zealand Journal of Educational Studies* (Pickens, 1976; 1979) provide some indication of the popularity and interest in research in this area over the past 30 years.

Investigations under this heading conducted by students at university since 1950 number 74. However, in keeping with my definition of research into 'educational measurement and testing', once the theses which are clearly 'psychological' (e.g., concerned with the measurement of personality via tests like the Rorschach, Rosenweig or 16 PF) are deleted, then the total number of student studies falls to something like 68. Of these, the majority are Diploma in Education studies (44); only two have been done by Diploma in Educational Psychology students. Twenty-one Masters studies have been completed: ten by Psychology students, ten by Education students and one by a student from a Department of Music! Some six theses on aspects of tests or testing have been presented at the University of Waikato for the B. Phil or M. Phil. degree. Only one study at doctoral level, that of Geoff Ord at Waikato, is listed under the heading Tests and Testing. Just one in 30 years! And it is questionable whether Dr Ord's work should be called New Zealand research.

Over the three decades the number of studies per decade has steadily increased from 20 in the 1950s to just over 30 during the 1970s. With increasing student numbers in Education departments this probably does not represent an increase in the popularity of aspects of testing as topics for theses, rather the reverse.

Interestingly, the majority of the studies in each decade have been conducted in the first five years with only a trickle of theses during the latter half. This was particularly so in the '50s and '70s.

It is tempting to speculate as to the reasons for the current paucity of student research in the area. Might it be reasonable to suggest that the contemporary disillusionment with certain kinds of testing and the many controversial issues surrounding the whole testing movement, particularly in the United States, might have something to do with the apparent lack of popularity? Only three theses have been completed since 1975 and presently only four studies are under way; two Dip. Eds., an M.Sc. in Psychology and an Education M.A. We can hardly argue that the topics for research in the field of educational measurement and testing are exhausted; rather it would appear that the 'climate' both within and without the universities is decidedly chilly towards testing at present.

Vague patterns emerge from analyses of what has been researched by students over the years 1950 to the present day. A cluster of ten Diploma studies, all done in the early '50s at the University of Auckland, are concerned with the standardization of Thurstone's *PMA*. These studies were all undertaken by students of Cyril Rogers and it appears that some kind of 'direction' was obviously operating at the time - or perhaps the students had caught the enthusiasm of their teacher?

Achievement testing has been the topic of a number of studies (approximately one-third) with mathematics a clear favourite. Somewhat surprisingly in a country obsessed with reading performance, only three studies over the whole time span have researched aspects of the measurement of reading skills; there have been more studies of music tests! Test construction or adaptation exercises are also fairly common, with standardization and validation studies rather less so.

None of the studies listed appear to investigate problems of educational measurement techniques. Indeed, no heading is provided for Educational Measurement in the *New Zealand Journal of Educational Studies* listings, nor is

there any other descriptor under which such research might be categorized. It is reasonable to state then, that this aspect of research has been virtually ignored by students for the 30 years under review. Perhaps it reflects a distaste for things mathematical or statistical in the typical Education and Psychology student? Or maybe there are more complex reasons? It is also evident that the majority, perhaps 99 percent, of the research conducted has been in cognitive aspects of the field. Yet I feel certain that a great many theses include as assessment tools, checklists, questionnaires, rating scales and similar measures which have been especially devised or adapted from existing measures. I also suspect that most are not pretested and are of unknown reliability. Nonetheless, this activity with instruments of this kind represents an unidentified facet of the research in educational measurement and testing.

As with the analysis of publications of university staff, I strongly suspect that the indications of the scope and the frequency of research into aspects of educational measurement and testing are not accurately represented by the listings of theses. Unless the descriptors 'test(s)' or 'testing' are identified in the thesis title, it will be categorized elsewhere. My own Masters thesis provides an excellent example. Listed under 'creativity', the study was much more the development of a series of valid and reliable measures to investigate creative thinking and the use of factor analysis to attempt to identify underlying abilities and provide evidence of construct validity. There must be many theses misclassified similarly. What is obviously required is multiple classification using a variety of descriptors, or better still, some New Zealand publication like *Dissertation Abstracts*. This is not to denigrate Keith Pickens' very useful work. I am certain that he would agree with me that detail or elaboration is warranted to allow for greater accessibility and use of research published in thesis form.

Department of Education Research

The most obvious source of information on research undertaken by officers of the Department of Education in educational measurement and testing is the entries under Educational Development/Research Activities in the *Annual Reports* presented to Parliament by the Department each year. An analysis of these reports from 1970 to the present indicates

that very little work has apparently been done in this area by Departmental officers. However, the reports of projects are abbreviated to the extent that it is difficult to ascertain whether the 'surveys', 'evaluations', 'pilot schemes', 'studies' and 'trials' actually contained a research component which could legitimately be termed 'educational measurement and testing'. There are tantalizing hints that experimentation and investigation have taken place; the difficulty is to identify the publication or report which may contain details. I have spent hours perusing Departmental publications such as the Curriculum Development Directorate bulletins and *Notes on Developments, Educational Development Conference* documents, Research and Statistics Division bulletins, School Certificate Examination Board reports, in-service course reports and the like. My best efforts have been frustrated; I have located few accounts of research that might fit my definition.

Since the early 1970s the Curriculum Development Directorate (formerly Unit) appears to have limited its activities in the field of testing and measurement to the construction and accompanying research of item banks, first in mathematics and more recently in science and foreign languages (Department of Education, 1978). There are indications that new assessment techniques for NESC English may also have been devised or investigated (Department of Education, 1976), but no reports of this projected research appear to have been published. Studies such as LARIC (Department of Education, 1978) and the Computation Survey (Department of Education, 1978), a longitudinal study, have elements of research into assessment methods and the development of measures specifically for the studies, but many of the evaluations of curriculum use established tests or non-test techniques.

The Department of Education's Examinations and Testing Unit, headed by Bruce Ramage, has, since the mid-70s, been actively developing and researching a reference test, known as the *Fifth Form Inter-school Moderating Test*, and to be used in moderating internally assessed School Certificate. A nationwide standardization of the measure has recently been completed and follow-up studies are in progress. A further moderating test for School Certificate Workshop Technology is being developed currently (Gavin, 1978; Ramage, 1979).

Another specialist group of Departmental officers in the IEA Unit has been extremely busy in preparing trial test materials, sampling designs and generally coordinating activity internationally for the massive second *IEA Mathematics Study*. Unlike previous IEA surveys, this study has a sizeable New Zealand input, and to my mind, particularly because of the research associated with the development of measures in both the cognitive and affective

domains, meets the criteria for inclusion in the definition of educational measurement being employed (Department of Education, 1979).

Departmental officers have engaged increasingly in collaborative research with staff from other educational institutions and it is hoped that some of this research might involve experimentation and innovations in the educational measurement and testing field. For example, an account of the *Forms 1-4 Science : Learning in Science : [Related to Age/Developmental Levels]* (Waikato University), indicates that assessment instruments and procedures are to be developed together with resource materials (Department of Education, 1978).

It is also probable, of course, that some Departmental research would be of a confidential nature and that reports of such work would not be available. I have it on good authority that this is so. Presumably then, we might expect that some of the Department of Education's findings or enquiries into aspects of educational measurement and testing have not been released for general consumption. There is nothing sinister or unethical in this; my point is simply that some research may have been undertaken of which researchers outside the Department are uninformed.

Research by Teachers in the Schools

If research in educational measurement and testing has been undertaken by teachers in the school setting it would seem appropriate for the magazines of the two major teachers' organizations to report accounts of such research from time to time. Yet, an analysis of *National Education* over nearly twenty years (1960-1978) and the *New Zealand PPTA Journal* over the same period reveals an astonishing lack of reports of research of any kind, let alone in the particular area of interest to us.

Using even the broadest of definitions of 'research' I could find no more than four in *National Education*, three of them on spelling (Smith and Pearce, 1966; Bodkin and Duncan, 1966; Croft, 1976) and another on the *Pacific Infant Performance Scale* written by a university researcher (A. St. George, 1973). Another, albeit satirical article, *Intelligence and the Pakeha Child*, by Archer *et al.*, (1971) is research based and should probably be considered as fitting my definition.

Five accounts appeared in the *New Zealand PPTA Journal*; one written by an Otago University staff member (Boyd, 1969), another by a psychologist (Silva, 1971), two were contributed by the same researcher (McCaw, 1974; 1976) on the same subject, the Napier Internal Assessment Scheme, and the other, an analysis of a multiple-choice test; a brief study that came about almost accidentally (Spence, 1966). On the other hand, there are numerous articles of a descriptive nature or pieces expressing opinions about examinations, tests, evaluation, assessment schemes, testing and the like; thousands of words written about the pros and cons of various kinds of testing or examining, in some cases quoting (and sometimes misquoting!) overseas research studies - but precious little in the way of accounts of New Zealand research on aspects of educational measurement and testing.

I understand that the Department of Education's magazine, *Education* maintained a policy which enabled accounts of research, particularly applied research conducted by teachers, to be published. Indeed, the current volumes have a special section towards the back of each number entitled 'Exchange', which seems to be admirably suited for just this purpose. A survey of articles and reports published in *Education* over the past decade which provide information, or accounts of research on topics of educational measurement or testing reveals that only two would fit the definition of research I am employing, both written by university staff members (McNaughton, 1974; Hughes, 1977). One other piece on the setting up of the Mathematics item banks contributed by an anonymous member of the Curriculum Development Unit might just warrant inclusion (1971).

Another line of enquiry into what research has been undertaken in this area leads to the publications: *Innovations, Experiments and Projects in Primary Education 1971* and *Innovations, Experiments and Projects in Schools and Colleges, 1974*, which are based on the reports received annually from the DSIs. But again, the entries/accounts/reports are few. One, possibly two, of some 95 entries in 1971 would fit my definition of research and of the six in the 1974 list of 92 entries, five are contributed by Curriculum Development Division personnel and one by a teachers' college lecturer - none by teachers.

However, as the introduction to each of these publications points out '... the projects described here are ... only a selection of those conducted...' or words to that effect, so we cannot know precisely how many studies or projects might have been undertaken by practitioners in the school

setting. Many of these informal school-based experiments or enquiries, I am certain, would fail to meet the stringent requirements of the professional or academic researcher. But, what we might call 'action research', grass-roots studies conducted to try to provide information that will help in making a decision about an immediate problem are undoubtedly done. Frequently, the findings will be confined to the particular school or at best to a group of local schools; they will seldom reach a wider audience, especially if there were no significant differences or the results were inconclusive, or, dare we say it, if they did not turn out in quite the way the researcher had hoped.

I thought originally that perhaps it was not that research is not being undertaken but that there was no vehicle, no way for teachers or others involved in the schools to have their studies and findings published. I am sure now that this is *not* the case; undoubtedly the outlets exist. Another possibility is that teachers are, for some reason, reluctant to use their professional magazines to publicize their work. But I cannot think why this would be so, unless, of course, they were submitting their articles to overseas journals, a course of action I believe to be unlikely.

From the published evidence available there would appear to be a dearth of teacher-conducted research. But the list of NZARE Personal Members gives at least four of the handful of teacher members as interested in the field of educational measurement and testing. I know personally that two of them have been actively engaged in investigating the suitability of tests of scholastic aptitude for the pupils in their school district and that written reports have been prepared (Forsyth, 1975). There are undoubtedly others. How do we find out about these people and what they have been doing?

Research in Teachers' Colleges

Staff of teachers' colleges have traditionally not been prolific contributors to the research in the area of tests and measurement. From my enquiries the names of only two researchers come to light: David Doake at Christchurch Teachers' College and his work on a

standardized academic aptitude test for college students in the early '70s and Murray Hulbert at Hamilton for his recent developmental work on fifth and sixth form moderating tests. One would hope that more research has been done in our teachers' colleges, but again it is extraordinarily difficult to locate the information even when appropriate sources have been identified. Happily, I understand that Dunedin Teachers College is to act as a clearinghouse for all teacher's college research in the future.

Research In the Armed Services

The Services have long used measures of aptitude and achievement for selection purposes. Almost without exception the approaches and tests employed would have been inherited from the British, at least up until the Second World War, and the majority of these would be categorized as 'psychological' rather than 'educational'.

With the beginning of World War II, personnel research came of age because of the pressures of selection, most notably for aircrew, but also for other specialists in the Armed Forces. Methods and measures from overseas - Canada, the UK and USA, were employed to assess requisite aptitudes, temperament and other personality variables. Psychologists shouldered the bulk of the load and names like Wrigley, Hearnshaw and Waite are associated with the work of this time.

After the war, programmes of testing for selection purposes were further developed and validated with psychologists working with educationalists such as Harry McQueen and Jim Caughley during the late 1940s. From that time, with many ups and downs, administrative shuffles and re-shuffles within the Services, a strong, coordinated Defence Psychology Unit has emerged whose work embraces occupational and, to a lesser extent, educational fields. Much of the research of the people employed in this unit is oriented towards improved selection procedures in which tests and testing play a substantial role. Reports of development, validation and similar studies are naturally confidential so that we may not know the true extent of the research undertaken but I am given to understand that it is relatively extensive and sophisticated. (See Toulson and Williams, 1979, for a detailed history of the work of the New Zealand Armed Forces psychologists).

Research in Technical Institutes

Some difficulty was experienced in identifying staff of technical institutes who had contributed to research in the field under scrutiny. I managed to locate only two studies by tutors : an interesting article by Pat Bruce (1974) on test difficulty in the *New Zealand Journal of Educational Studies* and another in *Technical Education* by Monty Mead (1977) on an analysis of students' examination marks and internally assessed performance. Undoubtedly, these studies represent only the tip of the iceberg. But where do polytechnic people publish their findings?

Research by Other Agencies

Vocational Guidance Service

Through the fifties and sixties a large number of *Research and Testing Notes*, some 89 in all, were produced by an extremely energetic Research and Testing Officer, Jack Jennings. These notes covered a wide range of tests and occupations. Many of them provide norms for occupational samples, others report test results and include interpretative comments, whilst others are comparative studies. Despite their brevity, these notes represent a very useful account of small-scale, applied research in a specialized corner of the field of educational measurement and testing. Regrettably, when Jennings left the Service in 1969 the research activity declined abruptly. Lack of staff, vacant senior positions and the pressures for other kinds of work meant that only a handful of investigations were conducted over the next nine years.

In April 1978, the Vocational Guidance Service was transferred officially from the Department of Education to the Department of Labour and a new appointment to the position of Research and Testing Officer was made in May. That appointee was Joe Williams, a psychologist with vast experience. Already Mr Williams has initiated programmes which will lead to the provision of updated norms for several of the tests which are still used by officers of the Service in their day-to-day work although for many the emphasis has shifted away from testing for the present.

Psychological Service

One branch of the Department of Education vitally concerned with testing is the School Psychological Service. While much of the work of the school psychologists is of a clinical nature, the assessment of school achievement and the measurement of scholastic aptitude are also part and parcel of many case studies.

In the 1960s two major research programmes in the field of testing were undertaken. In 1963, Rogers tackled a New Zealand standardization of Raven's *Coloured Progressive Matrices* (Rogers, 1963) and the *Revised Tomlinson Junior School Test* was standardized nationally in 1964 (Callander *et al.* 1967). Currently, preparatory work is underway for another national standardization; this time of the *WISC-R*. Two other major enquiries in progress: one by Coleman on criterion-referenced assessment and another under the leadership of Eddy Eggers, at Palmerston North, on the *Burt Word Reading Test*.

Many small-scale studies on aspects of testing, predominantly investigations of the suitability of new tests for local conditions, are conducted by Psychological Service personnel. Examples of studies of this nature would be Peter du Chateau's work with the *Peabody Picture Vocabulary Test*, Craig Jackson's investigation of the *Aston Index*, R.W.K. Dawson's work with Thurstone's *PMA* in the Hutt Valley - Wellington region and the Christchurch Psychological Services' development of the *C-R Word Recognition Test*. In the main these studies are not funded centrally, or at all, and rely on the energy and enthusiasm of the individual psychologist for their conduct and completion. Findings would be known to local colleagues and may be disseminated more widely, depending on their perceived importance and impact. Again, it is difficult to specify just how numerous studies of this type are, but from my own personal knowledge some psychologists are quite actively involved in enquiries of this nature.

New Zealand Council for Educational Research

Since its inception in 1934 the Council has had an association with research in educational measurement and testing. In the early years this was largely through sponsorship or support of honorary research workers investigating aspects of the field and the publication of research monographs, e.g., *Standardizing Shorthand Tests* (James, 1944); *The Stanford-Binet Scale : its Suitability for New Zealand* (Fitt, 1952).

The exception to this pattern would be Redmond and Davies' standardization of the *Otis Self-Administering Tests of Mental Ability - Intermediate Form* and the *ACER : Non-verbal Test* and the anonymous standardization of *The Metropolitan Readiness Tests* in 1942 conducted by members of the Council's staff. Some adaptation of overseas' tests has also been undertaken by staff working on particular projects, e.g., the New Zealand revision of the *Minnesota Teacher Attitude Inventory* and several achievement tests of overseas' origin (Watson, 1958). Certainly, NZCER was the principal focus of the small amount of test development, as opposed to norming activity that had gone on. Explaining the lack of involvement in this sphere, Parkyn (NZCER, 1965), in a reiteration of Campbell's view of 20 years before, stated 'NZCER has been interested in tests mainly as an adjunct to its wider research activities. This is the main reason why it has avoided committing a large part of its slender resources to a more comprehensive development of educational or psychological tests.' The Council at this time showed a distinct disinclination to have tests and testing dominate its activities.

With the establishment of the Test Development Division in 1965 (following the recommendations of the Currie Commission quoted earlier and still further recommendations made to the Minister of Education by a special three-man committee) and the appointment late in 1965 of Warwick Elley as its energetic and talented head, test construction activities and related research studies moved into high gear. The Division had as its brief: 'The main function of the Division would be to develop tests and keep them under continuous review. Much of its early work would be directly related to the construction of those tests of attainment recommended by the Commission for primary schools, but it would not be limited to such work' (NZCER, 1965). Some preparatory work had, in fact, already been done by NZCER in commissioning technical briefs in subject areas (Fieldhouse, Freyberg and Lenz) and in general intellectual ability (Arvidson) prior to its acceptance of the Government's proposal.

Since the mid-sixties, NZCER's small Test Development Division has produced achievement tests for primary and secondary pupils in Reading Vocabulary (1969), Reading Comprehension (1969), Listening

Comprehension (1971), Mathematics (1974) and Study Skills (1978) in its *Progressive Achievement Test* series and re-normed the *Otis Tests of Mental Ability (Intermediate Form)* concurrently with the tests of reading (Elley, 1969). The development of these tests is fully documented in Chapter 6 of each of the teacher's manuals which accompany the tests.

The construction of each new test presents different problems which are researched by staff. Such mundane matters as print size, column width, page format, use of answer sheets vs answering in the booklet, speed of reading, methods of answering, content of directions, placement of stimulus material, use of colour and the like are investigated and mentioned or briefly outlined under Concurrent Research in the manuals for the tests. In some cases the studies are documented or written up in report form for limited dissemination and a few warrant publication, e.g., *The Effects of Coloured Stimulus Material on Study Skills Achievement* (Reid, Croft and Jackson, 1977); *Answer Changing on Objective Tests* (Jackson, 1978).

The validation exercises undertaken as part of the construction phase are also published in the teacher's manuals as are all the details of the actual development and standardization of the measures. Occasionally some of these studies may also be written up as research articles, e.g., *A Factorial Analysis of the PAT : Reading Comprehension, Reading Vocabulary and Listening Comprehension Tests* (Reid and Hughes, 1974); *A Rasch Analysis of a Test from the Progressive Achievement Test Series* (Jackson, 1979).

In line with the recommendation that the 'tests be kept under continuous review', investigations with a view to re-norming are conducted at regular intervals and the findings reported (Reid, 1975; Reid, 1978). Currently an interim report for a re-norming study of the *PAT : Mathematics Tests* is in preparation.

As Parkyn stated in 1965, 'the work of the Division will go on to the development of tests which will be used for a variety of purposes in the educational services and industry and commerce'. This has been the case increasingly over recent years with several major test development projects being undertaken at one time in addition to the 'bread and butter' work associated with the development and maintenance of the *Progressive Achievement Test* series.

Quite apart from the assistance given to NZCER honorary research workers, other NZCER staff, Government agencies, schools and other teaching institutions and commercial enterprises in developing their own educational tests and measures and/or using them to better purpose, officers of the Test Development

Division have been responsible for constructing and validating the *Language Achievement Test for Overseas Studies* (LATOS) for the Universities Grants Committee, and the *Police Entrance Test* for the New Zealand Police Department. The construction of two other tests of a similar type for possible development in 1980 are being discussed at present.

Other measures have also been developed, most notably, the *Study Habits Evaluation and Instruction Kit* (SHEIK) (Jackson, Reid and Croft, 1979), and *FOSSIC*, a rating scale for teachers to use in observing pupils' study behaviour (Jackson, Reid, Croft and Cowie, 1979). Both measures were designed to complement the *PAT : Study Skills Tests*, and were thoroughly researched.

Presently in the pipeline are several other tests or standardizations of overseas measures. A scholastic aptitude test for children aged 9 to 14 years, the first of two (the second a multi-factorial model), is to be standardized in April 1980. A spelling test of the proof-reading kind will be standardized concurrently. Plans are also afoot for New Zealand standardizations of Raven's *Standard Progressive Matrices*, the *Burt Word Reading Test*, and several advanced tests of general ability - *ACER Higher Tests AL-AQ, BL-BQ; B40/78*, and the construction of a measure of study attitudes to round out the SHEIK is also being investigated.

As with other institutional research of less-than-major importance, a good deal of what is investigated by Test Development Division staff remains unpublished and known only to a small coterie of researchers who have been involved, usually in a consultative role, at some stage during the research project.

Oversea Influences

In my historical overview I remarked that despite this country's remoteness and isolation from world centres of test development and research we have managed to remain tolerably well informed of overseas developments.

Today this information reaches us largely through articles in scholarly journals, reports disseminated by large research organizations (ETS, OISE, NFER, etc.) and through active participation or observation by New Zealand researchers on study or training leave or on exchange with other institutions.

To some degree we are always a little behind the play, particularly in psychometrics and to a lesser extent in test development work. The 'trickle theory' is clearly in operation; research appearing in journals is often years old if the time from the submission of the manuscript to the arrival of the journal on the library shelves is calculated. Sometimes it is also necessary to wait for sophisticated technology, and for trained personnel to handle the hardware to become available. Hence, we have come rather late to 'new' developments such as item banks, mastery learning and testing, criterion-referenced assessment, minimal competency tests, Rasch analysis, multiple-matrix sampling and the like. In some respects this is by no means a bad thing. The Americans, Australians, Canadians and sometimes the British encounter all the inevitable teething troubles and we occasionally learn from their mistakes and excesses. I would like to believe that our belated arrival on the scene sometimes permits us to do the job better. A good example of this is Warwick Elley's development of content-referenced norms for the *PAT : Reading Tests*; a breakthrough of sorts.

Presently, the major influences are from North America and in some specific areas of research from ACER in Australia and NFER in England. Such leadership is to be anticipated given the resources in terms of personnel and money that these countries are able to command and the opportunity they have to specialize. Whether they do a better job, dollar for dollar, is another matter. In this country we are severely limited in a number of ways and for the moment most of us must remain part-time researchers and in true New Zealand style, supreme generalists - able to turn a hand to most things.

Research Trends and Emphases

One facet of research from the 'evaluation of suitability' mould, and one which cuts across institutional boundaries, has been the relatively numerous studies of the validity of tests, notably scholastic aptitude and achievement measures, for Maori children. Living, as we do, in a multi-cultural society, this concern for the appropriateness of these tests can be expected. Although many of the studies are essentially comparative exercises, a proportion also contain an investigation of the suitability of the tests. St. George (1977) provides an excellent and informative summary of these research studies spanning 50 years or so in his doctoral dissertation. He notes (i) the exclusion of Maori children from most of the early standardizations and (ii) the lack of separate Maori and European analyses. St. George recommends

that, in line with APA standards, future researchers should investigate matters of test reliability, test validity and test item characteristics separately for minority populations.

Historically it is possible, despite the meagre number of studies and the fragmented nature of the field, to detect several phases in the research on educational measurement and testing. The earliest was an experimental phase - overseas tests were tried and their suitability for New Zealand conditions was evaluated. Local norms were developed for those judged to be satisfactory. To some extent, this trend has persisted to the present and will no doubt continue in the future. For the moment it is left to the enthusiasm of a handful of researchers to investigate tests and measures released recently by overseas publishers and institutions (e.g., *British Ability Scales*, *WISC-R*, *Tests of Children's Learning Ability*, *Tests of English for Migrant Students*).

Then came a period of national standardizations in the areas of general mental ability, reading and arithmetic. With one exception, these were unadapted overseas measures for which New Zealand norms were derived.

With the growth of a greater research capability within the Department of Education together with the establishment of specialist divisions and the setting up of the Test Development Division at NZCER, tests or assessment devices designed to fulfil particular functions have been developed specifically for New Zealand conditions. While the range and variety of these measures is by no means extensive, teachers and others are, I believe, better served than they have ever been in the past.

The vast majority of the research has focussed on cognitive aspects within the field; the affective and psychomotor domains remain virtually untouched. So, too, the area of psychometrics where those few investigations that have been mounted have been initiated by 'professional' researchers or university staff specializing in statistics/psychometrics.

Conclusion

From the fragmented and obviously incomplete picture I have presented it is difficult to reach any other conclusion, pedestrian and unimaginative as it may be, that more and better research is necessary in every facet of the field. There is no lack of pressing problems, both for the

practitioner and for the policy-maker. It is not within the scope of this paper to state what these problems are or who is to research them - that is quite a different story.

I have been rather optimistic in reviewing the research on educational measurement and testing; I want desperately to believe that what is published in journals and reports represents only a fraction of what is actually being undertaken in this country. Perhaps I am deluding myself; but there are faint signs and intimations as I have indicated, and informants in a variety of institutions approached as part of this exercise in presenting the state of the art seem to share my impression.

What is abundantly clear, I think, is the lack of documentation, publication and communication among the researchers working in and around the field of educational measurement and testing. It transpires that few of us know what others are doing now, or have done recently. Obviously, too, what research activity there is can be of little value unless available information can be put into the hands of those practitioners and administrators to whom our findings might be of use. Therein lies the challenge; perhaps the fledgling NZARE may be able to provide the forum and the dissemination network whereby we achieve a free, fast flow of information to those who seek it or need it.

References

- Archer, D., Oppenheim, R.S., Karetu, T.S. and St. George, R. (1971) Intelligence and the Pakeha Child. *National Education*, 53, 577, 258-60.
- Arvidson, G.L. (1959) *The Hutt Valley F.II. Classification Scheme*. Wellington: Education Department, Victoria University of Wellington. (Mimeo)
- Arvidson, G.L. (1974) Personal Communication.
- Bodkin, J.P. and Duncan, P.A. (1966) TNI Test with New Zealand. Norms and Parallel Version. *National Education*, 48, 519, 121-24.
- Boyd, R.M. (1969) University Entrance English : An Analysis of the 1966 Examination Paper. *The New Zealand PPTA Journal*, 16, 1, 9-10.
- Bradshaw, J. (1978) *The Dunedin Reading Test*. Unpublished paper.
- Bruce, P.H. (1974) An Index of Test Difficulty, with Applications. *New Zealand Journal of Educational Studies*, 9, 1, 31-41.
- Callander, D.J., Marsh, R.W. and Milne, K.A. (1967) *Revised Tomlinson Junior School Test*. Wellington: Department of Education.
- Clay, M.M. (1972) *Sand-Concepts about Print*. Auckland: Heinemann Educational Books.
- Clay, M.M., Gill, M., Glynn, E., McNaughton, A.H. and Salmon, K. (1976) *Record of Oral Language*. Wellington: New Zealand Educational Institute.
- Croft, A.C.C. (1976) Spelling Levels : Are They Still Being Used? *National Education*, 58, 622, 179-81.
- Department of Education. (1971) *Innovations, Experiments and Projects in Primary Education*. Wellington: Government Printer.
- Department of Education. (1971) *Mastery in Mathematics*. *Education*, 20, 6, 9-12. Wellington: Government Printer.
- Department of Education. (1974) *Innovations, Experiments and Projects in Schools and Colleges*. Wellington: Government Printer.
- Department of Education. (1976) *Notes on Developments*. Wellington: Curriculum Development Unit. (Mimeo)
- Department of Education. (1978) *Educational Standards in State Schools*. Wellington: Government Printer.

- Department of Education. (1978) *Notes on Developments*. Wellington: Curriculum Development Directorate. (Mimeo)
- Department of Education. (1979) *Research and Statistics Division : Bulletin No. 6*. Wellington: Government Printer.
- Elley, W.B. (1969) Changes in Mental Ability in New Zealand School Children, 1936-1968. *New Zealand Journal of Educational Studies*, 4, 2, 140-55.
- Elley, W.B. and Reid, N.A. (1969) *Progressive Achievement Tests of Reading Comprehension and Reading Vocabulary*. Wellington: New Zealand Council for Educational Research.
- Elley, W.B. and Reid, N.A. (1971) *Progressive Achievement Tests of Listening Comprehension*. Wellington: New Zealand Council for Educational Research.
- Fieldhouse, A.E. (1952) *Oral Word Reading Test*. Wellington: New Zealand Council for Educational Research.
- Fitt, A.B. (1952) *The Stanford-Binet Scale. Its Suitability for New Zealand*. Wellington: New Zealand Council for Educational Research.
- Forsyth, A.B. (1975) *An Analysis of Trends in School Rolls and Public Examination Passes for the Years 1960 onwards in Comparison with Apparent Trends in Intelligence Test Scores*. (Unpublished report)
- Gavin, A.P. (1978) *Report to the Advisory Committee on Educational Tests*. (Unpublished report)
- Hughes, D.C. (1977) Marking the Progressive Achievement Tests. *Education*, 26, 3, 22-4.
- Hughes, D.C. and Tuck, B.F. (1978) The Stability of Scores and Difference Scores in the Progressive Achievement Tests of Listening and Reading Comprehension. *New Zealand Journal of Educational Studies*, 13, 1, 67-77.
- Jackson, P.F. (1978) Answer Changing on Objective Tests. *The Journal of Educational Research*, 71, 6, 313-15.
- Jackson, P.F. (1979) A Rasch Analysis of a Test from the Progressive Achievement Test series. *New Zealand Journal of Educational Studies*, 14, 1, 67-73.
- Jackson, P.F., Reid, N.A. and Croft, A.C.C. (1979) *The Study Habits Evaluation and Instruction Kit (SHEIK)*. Wellington: New Zealand Council for Educational Research.
- Jackson, P.F., Reid, N.A., Croft, A.C.C. and Cowie, C. (1979) FOSSIC : A Study Skills Rating Schedule. *set*, No. 1, 1979, Item 3.

- James, T.L. (1944) *Standardizing Shorthand Tests*. Wellington: New Zealand Council for Educational Research.
- Kay, G.A. (1978) *An Investigation into the Implementation of Internal Assessment for School Certificate Mathematics Under the Canterbury-Westland Scheme*. (Mimeo)
- Keeling, B. and Seddon, R. (1969) A Test Battery for Form III Entrance Classification. *New Zealand Journal of Educational Studies*, 4, 1, 57-68.
- McCaw, B. (1974) Internal Assessment and the Problems of Moderation. *New Zealand PPTA Journal*, May, 33-7.
- McCaw, B. (1976) The Napier Internal Assessment Scheme. *New Zealand PPTA Journal*, November, 21-3.
- McIntyre, G.A. (1938) *The Standardization of Intelligence Tests in Australia*. Melbourne: Australian Council for Educational Research.
- McKenzie, N.R. (1930) *School Surveys, 1927 and 1929*. New Plymouth: North and South Taranaki branches of the New Zealand Educational Institute.
- McNaughton, A.H. (1974) *Social Studies Inference Test, Std. 2 - Form 2*. Auckland: Government Printer.
- McNaughton, A.H. (1974) The Development of a Primary School Social Studies Test : A Cooperative Enterprise. *Education*, 23, 6, 23-6.
- Marsh, R.W. (1969) An Objective Method of Defining the Region of the Pass/Fail Cutting Point in University Examinations. *British Journal of Educational Psychology*, 39, 3, 305-8.
- Mead, M. (1977) What's the Difference? *Technical Education*, 1, 1, 14-15.
- New Zealand Council for Educational Research. (1942) *The Metropolitan Readiness Tests*. Wellington: New Zealand Council for Educational Research.
- New Zealand Council for Educational Research. (1948) *Thirteenth Annual Report*, 15-16. Wellington: New Zealand Council for Educational Research.
- New Zealand Council for Educational Research. (1956) *ACER : Silent Reading Tests*. Wellington: New Zealand Council for Educational Research.
- New Zealand Council for Educational Research. (1956) *ACER : Arithmetic Tests*. Wellington: New Zealand Council for Educational Research.

- New Zealand Council for Educational Research. (1965) *Thirtieth Annual Report*, 12-21. Wellington: New Zealand Council for Educational Research.
- New Zealand Council for Educational Research. (1972) *Thirty-seventh Annual Report*, 54-5. Wellington: New Zealand Council for Educational Research.
- Nicholls, J.G. (1971) Some Effects of Testing Procedure on Divergent Thinking. *Child Development*, 42, 5, 1647-51.
- Phillipps, W.R. (1976) I.E.A. - International Standards of Achievement. *Education*, 25, 6, 10-11.
- Pickens, K.A. (1976) Unpublished Degree and Diploma Studies in Education. *New Zealand Journal of Educational Studies*, 11, 2, 191-217.
- Pickens, K.A. (1979) Unpublished Degree and Diploma Studies in Education. *New Zealand Journal of Educational Studies*, 14, 1, 76-95.
- Ramage, B.B. (1979) *Report to the Advisory Committee on Educational Tests*. (Unpublished report)
- Redmond, M. and Davies, F.R.J. (1940) *The Standardization of Two Intelligence Tests*. Wellington: New Zealand Council for Educational Research.
- Reid, N.A. (1975) Do the PAT : Reading Tests Require Re-Norming? *New Zealand Council for Educational Research Newsletter*, 7, 1, 2-3.
- Reid, N.A. (1978) Listening Comprehension : 7-Year Follow Up. *New Zealand Council for Educational Research Newsletter*, 10, 1, 2.
- Reid, N.A. and Hughes, D.C. (1974) *Progressive Achievement Tests of Mathematics*. Wellington: New Zealand Council for Educational Research.
- Reid, N.A. and Hughes, D.C. (1974) A Factorial Analysis of the PAT : Reading Comprehension, Reading Vocabulary and Listening Comprehension Tests. *New Zealand Journal of Educational Studies*, 9, 1, 18-30.
- Reid, N.A., Croft, A.C.C. and Jackson, P.F. (1977) The Effects of Coloured Stimulus Material on Study Skills Achievement. *New Zealand Journal of Educational Studies*, 12, 1, 66-71.
- Reid, N.A., Croft, A.C.C. and Jackson, P.F. (1978) *Progressive Achievement Tests of Study Skills*. Wellington: New Zealand Council for Educational Research.
- Rogers, C.A. (1956) *Measuring Intelligence in New Zealand*. Auckland: Auckland University College.

- Rogers, D.G. (1963) *Raven's Coloured Progressive Matrices (Sets A, Ab, B) - New Zealand Standardization*. Christchurch: Department of Education Psychological Service. (Mimeo)
- St. George, A. (1973) Towards a Sounder Basis for Testing. *National Education*, 55, 597, 115-16.
- St. George, W.V.R. (1977) *The General Cognitive Ability of Maori and European Subjects : A Comparative Study of the Queensland Test Performance and Psychometric Properties*. Unpublished D. Phil. Thesis, University of Waikato.
- Shouksmith, G.A. (1960) *The Canterbury Reasoning Test*. Christchurch: Whitcombe and Tombs.
- Shouksmith, G.A. (1964) *The Tasman Junior Reasoning Tests ; The Tasman Verbal Ability Test*. Christchurch: Whitcombe and Tombs.
- Silva, P.A. (1971) Teachers' Children and the Otis Test. *New Zealand PPTA Journal*, July, 37-9.
- Silva, P.A. (1978) *A Study of the Concurrent and Predictive Validity of the Reynell Developmental Language Scale*. Dunedin: University of Otago. (Mimeo)
- Smith, C.T.W. and Pearce, D.W. (1966) Attainment Norms and Comparison for Pupils from 9-13 years. *National Education*, 48, 519, 117-121.
- Spence, R.J. (1966) Multi-Choice Questions in Mathematics Exams : A Statistical Enquiry. *New Zealand PPTA Journal*, 8, 10, 6.
- Toulson, P.K. and Williams, W.C.J. (1979) History of Personnel Research and Psychological Services in the New Zealand Armed Forces. *Delta*, Monograph No. 2, 77-84.
- Tuck, B.F., Hanson, A.L. and Zimmerman, M. (1975) The WISC-R : A New Zealand Study of Norms and Validity. *New Zealand Journal of Educational Studies*, 10, 1, 52-8.
- Watson, J.E. (1958) *The Attainment of Teachers' College Students in Primary School Subjects*. Bulletin No. 7. Wellington: New Zealand Council for Educational Research.
- Watson, J.E. (1979) *Fifty Years of Test Supply - The New Zealand Experience*. Paper presented at the 50th Jubilee Conference, New Zealand Psychological Society, Massey University, Palmerston North, August, 1979.